



**HAL**  
open science

## Subunit and chlorophyll organization of the plant photosystem II supercomplex

Laura S. van Bezouwen, Stefano Caffarri, Ravindra S. Kale, Roman Kouřil, Andy-Mark W.H. Thunnissen, Gert T. Oostergetel, Egbert J. Boekema

► **To cite this version:**

Laura S. van Bezouwen, Stefano Caffarri, Ravindra S. Kale, Roman Kouřil, Andy-Mark W.H. Thunnissen, et al.. Subunit and chlorophyll organization of the plant photosystem II supercomplex. *Nature Plants*, 2017, 3, pp.17080. 10.1038/nplants.2017.80 . cea-01896536

**HAL Id: cea-01896536**

**<https://cea.hal.science/cea-01896536>**

Submitted on 28 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Subunit and chlorophyll organization of the plant photosystem II supercomplex

Laura S. van Bezouwen<sup>1†</sup>, Stefano Caffarri<sup>2</sup>, Ravindra S. Kale<sup>3</sup>, Roman Kouřil<sup>3</sup>,  
Andy-Mark W. H. Thunnissen<sup>1</sup>, Gert T. Oostergetel<sup>1</sup> and Egbert J. Boekema<sup>1\*</sup>

Photosystem II (PSII) is a light-driven protein, involved in the primary reactions of photosynthesis. In plant photosynthetic membranes PSII forms large multisubunit supercomplexes, containing a dimeric core and up to four light-harvesting complexes (LHCs), which act as antenna proteins. Here we solved a three-dimensional (3D) structure of the C<sub>2</sub>S<sub>2</sub>M<sub>2</sub> supercomplex from *Arabidopsis thaliana* using cryo-transmission electron microscopy (cryo-EM) and single-particle analysis at an overall resolution of 5.3 Å. Using a combination of homology modelling and restrained refinement against the cryo-EM map, it was possible to model atomic structures for all antenna complexes and almost all core subunits. We located all 35 chlorophylls of the core region based on the cyanobacterial PSII structure, whose positioning is highly conserved, as well as all the chlorophylls of the LHCII S and M trimers. A total of 13 and 9 chlorophylls were identified in CP26 and CP24, respectively. Energy flow from LHC complexes to the PSII reaction centre is proposed to follow preferential pathways: CP26 and CP29 directly transfer to the core using several routes for efficient transfer; the S trimer is directly connected to CP43 and the M trimer can efficiently transfer energy to the core through CP29 and the S trimer.

The photosynthetic reactions in plants, cyanobacteria and algae are catalysed by four major protein complexes, photosystem I (PSI), PSII, cytochrome *b<sub>6</sub>f* and ATPase. These proteins are embedded in the thylakoid membranes, and in plants these membranes are located in a special organelle, the chloroplast. The primary photosynthetic reactions are light driven, and the harvested energy is used by PSI and PSII to transport electrons over the membranes and to establish a membrane proton gradient<sup>1,2</sup>.

To understand the photosynthetic reactions, detailed knowledge about the structure of these protein complexes is necessary. For plant PSI high-resolution structures are known<sup>3</sup>, but for plant PSII the structural information is limited. For cyanobacteria<sup>4–6</sup> and red algae<sup>7</sup> there is a high-resolution PSII structure based on X-ray crystallography, but no high-resolution structure for higher plants could be obtained by protein crystallography. Very recently this goal was achieved by single-particle cryo-EM and image processing<sup>8</sup>.

PSII is a large multisubunit protein complex containing a dimeric core and a number of peripheral membrane-embedded antenna complexes. The core complex comprises between 20 and 23 protein subunits, depending on the organism. The catalytic heart of the core is the reaction centre, which is highly conserved between plants, algae and cyanobacteria. The reaction centre consists of four subunits, PsaA (D1), PsaB (CP47), PsaC (CP43) and PsaD (D2), which are also the largest membrane-intrinsic subunits. PsaA and PsaD form the photochemical reaction centre where the charge separation takes place as well as electron transfer over the membrane. Both subunits bind in total six chlorophylls (Chls), and PsaA has in addition two pheophytins. PsaB and PsaC are the internal antenna proteins, which bind several chlorophylls. These two subunits are involved in light harvesting and transporting excitation energy from peripheral antenna subunits towards the photochemical reaction centre<sup>1</sup>. In addition

to the reaction centre there are several other small intrinsic subunits, which are present in all organisms: PsaE, PsaF, PsaH, PsaI-M, PsaTc, PsaX, PsaY and PsaZ. These subunits are structurally and functionally conserved but less strongly related between organisms than those of the reaction centre. Plant PsaY has, for instance, an additional membrane spanning helix as compared to its bacterial counterpart. Plants lack the cyanobacterial subunit Ycf12 but have two additional subunits, PsaTn and PsaW. Their location could be established in the plant supercomplex structure<sup>8</sup>.

PSII has a special set of three to four extrinsic subunits of the oxygen-evolving complex (OEC) which shields the water splitting machinery<sup>9</sup>. The 33 kDa subunit PsaO, which stabilizes the manganese complex, is present in all organisms. The set of smaller subunits is variable. Plants have PsaP, PsaQ and PsaR<sup>10</sup> of which PsaP and PsaQ subunits are involved in optimizing the oxygen evolution at physiological concentrations of calcium and chloride<sup>9</sup>. PsaR needs PsaJ for stabilization and seems important for the assembly of PsaP<sup>9</sup>. The location of the extrinsic subunits for plants is now largely established<sup>8</sup>.

The dimeric plant core complex (C<sub>2</sub>) can form supercomplexes with up to six LHCII trimers, which are the major components of the antenna system. LHCII consists of heterotrimers composed of different combinations of Lhcb1, Lhcb2 and Lhcb3<sup>11</sup>. These trimers are linked to the core complex by the minor monomeric antenna proteins Lhcb4 (CP29), Lhcb5 (CP26) and Lhcb6 (CP24)<sup>12</sup>. At present, the largest purifiable PSII supercomplex, the C<sub>2</sub>S<sub>2</sub>M<sub>2</sub> particle, binds four LHCII trimers, two strongly bound (S) and two moderately strongly bound (M) trimers. Two additional, loosely bound (L), trimers could be detected in spinach supercomplexes (indicated as C<sub>2</sub>S<sub>2</sub>M<sub>2</sub>L<sub>2</sub>) and *A. thaliana* supercomplexes and megacomplexes<sup>13</sup>.

The light-harvesting proteins have been the subject of many studies. Over the years several crystal structures were solved for

<sup>1</sup>Electron microscopy and Protein crystallography group, Groningen Biomolecular Sciences and Biotechnology Institute, University of Groningen, 9747 AG Groningen, The Netherlands. <sup>2</sup>Aix Marseille Université, CEA, CNRS, BIAM, Laboratoire de Génétique et Biophysique des Plantes, 13009 Marseille, France. <sup>3</sup>Centre of the Region Haná for Biotechnological and Agricultural Research, Department of Biophysics, Faculty of Science, Palacký University, Šlechtitelů 27, 783 71 Olomouc, Czech Republic. <sup>†</sup>Present address: Cryo-Electron Microscopy, Bijvoet Centre for Biomolecular Research, Faculty of Science, Utrecht University, 3584 CH Utrecht, Netherlands (L.S.v.B.). \*e-mail: [e.j.boekema@rug.nl](mailto:e.j.boekema@rug.nl)

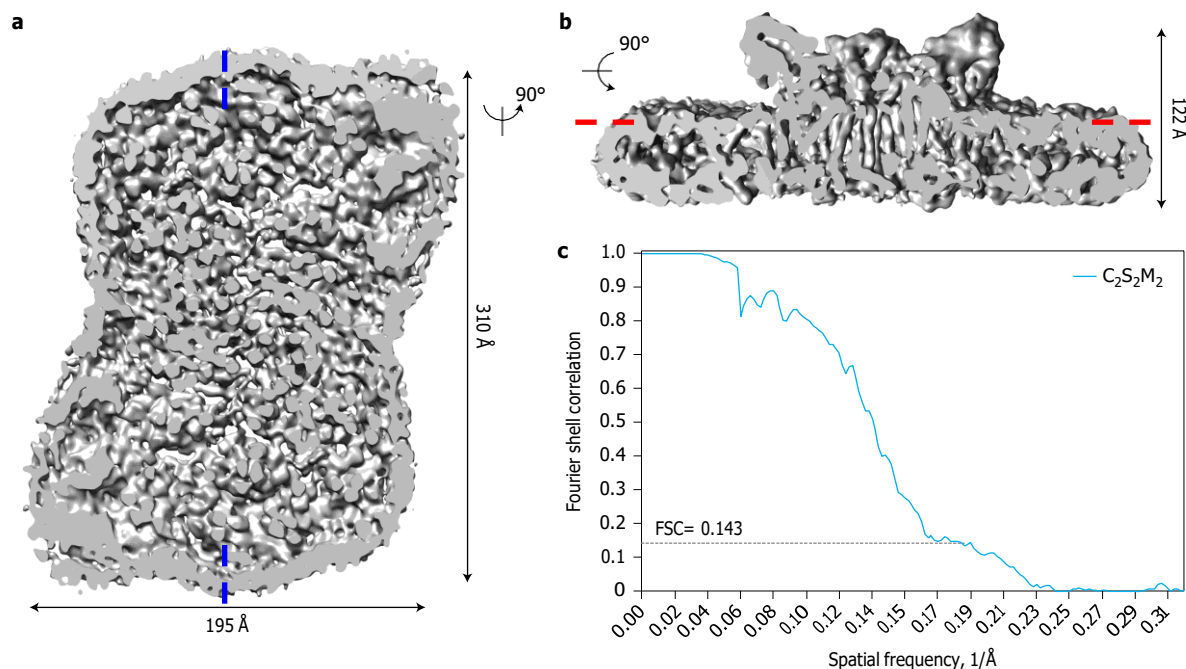


Figure 1 | Architecture of the  $C_2S_2M_2$  particle at an average resolution of 5.3 Å. a, Cryo-EM density map of the final supercomplex. Blue lines indicate where the protein was sliced as shown in b. b, Density inside the complex shown vertical to the membrane plane. Red lines indicate where the protein was sliced as shown in a. c, FSC curve of two independently refined half datasets of the  $C_2S_2M_2$  supercomplex (blue). The resolution based on the gold-standard criterion (0.143) is 5.3 Å.

both the LHCII trimers from spinach and pea<sup>14,15</sup> and for the monomeric CP29 from spinach<sup>16</sup>. The monomeric LHCII subunit and CP29 show significant homology in amino acid sequence and overall structure, but there are some relevant differences<sup>17</sup>, in particular with respect to carotenoid and chlorophyll composition. The monomeric antenna CP29 binds three carotenoids (one lutein, one violaxanthin and one neoxanthin), and LHCII binds four carotenoids per monomer (two luteins and one neoxanthin as in CP29, plus one violaxanthin/lutein in an additional peripheral mixed site). Each LHCII monomer contains 14 chlorophylls, of which eight are Chl *a* and six are Chl *b*. On the other hand, CP29 has 13 chlorophylls, of which eight are Chl *a* and four are Chl *b* with a mixed Chl *a/b* site present at binding site 610. In addition to a variable number of chlorophylls there is a difference in their location within the antenna proteins. In comparison to a LHCII monomer, CP29 is lacking chlorophylls at binding sites 601 and 605. Conversely, in CP29 a chlorophyll is present at binding site 615, which is absent in the LHCII monomers<sup>17</sup>. Monomeric CP26 and CP24 are predicted to contain 13 and 10/11 chlorophylls, respectively, based on sequence homology and biochemical evidence<sup>18</sup>. In the absence of crystal structures of CP26 and CP24, it is unclear which chlorophylls are present, and what is their exact location and orientation in these antenna proteins.

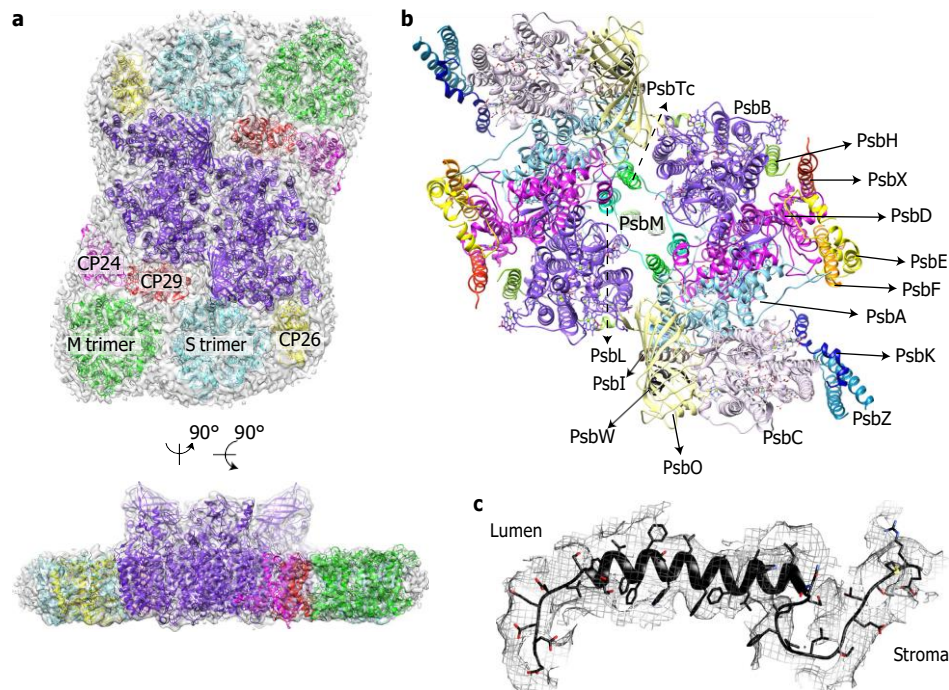
To fully understand the flow of excitation energy within plant PSII, it is crucial to have a high-resolution 3D structure of a PSII supercomplex. Because of recent developments in electron detectors and in image processing<sup>19</sup>, single-particle electron microscopy is nowadays able to solve structures at near-atomic resolution. The spinach  $C_2S_2$  supercomplex, lacking a large part of the antenna system, was the first plant PSII supercomplex to be determined at high resolution<sup>8</sup>. The largest PSII supercomplex, the  $C_2S_2M_2$ , is more abundant in *Arabidopsis*, but only a negative stain projection map at a 12 Å resolution is available<sup>20</sup>. Based on the new electron microscopy developments we managed to obtain a 3D cryo-EM map of the  $C_2S_2M_2$  supercomplex from *Arabidopsis* at an overall resolution of 5.3 Å, allowing reconstruction of its 3D structure.

The structure contains all the membrane-bound core subunits conserved in different PSII complexes, except PsbJ and PsbY, as well as the LHCII trimers and minor antenna complexes. The supramolecular organization of the subunits in the  $C_2S_2M_2$  is described in particular with respect to the location and orientation of LHCII trimers and minor antenna complexes. We also discuss the location of all the chlorophylls in the subunits and their role in the energy flow from the peripheral antennae to the core.

## Results

To gain an insight in the plant PSII function we determined the structure of a  $C_2S_2M_2$  particle by single-particle cryo-EM. For this purpose, a large dataset with movie frames was collected (Supplementary Fig. 1) and processed (Supplementary Figs 2 and 3). We used particles purified from a *npq4* mutant, lacking PsbS which was found to be present in sub-stoichiometric amount in  $C_2S_2M_2$  particles purified from wild-type membranes<sup>20,21</sup>. Though high purity is not a prerequisite, it is considered to be advantageous in 3D single-particle analysis. The final structure with a twofold symmetry imposed has an overall resolution of 5.3 Å (Fig. 1). At this resolution the transmembrane helices of the subunits are resolved and therefore the location of all the different subunits could be determined. The overall resolution is 5.3 Å, but there is a large variance within the map (Supplementary Fig. 4). The core region is well determined, with a local resolution between 4 Å and 5 Å, but the resolution in the periphery of the complex, where the antenna proteins are located, is not as high as in the centre.

To allow a reliable reconstruction of the 3D structure of the  $C_2S_2M_2$  particle, the cryo-EM map was divided into different regions of more or less uniform resolution. Three regions were identified in the cryo-EM map, encompassing the core, the S trimers with subunits CP26 and CP29, and the M trimers with subunits CP24. First, a 3D structure of the core region was built and fitted into the cryo-EM map, using the 1.9 Å cyanobacterial PSII structure<sup>4</sup> as a starting model. The core of cyanobacterial PSII, without the extrinsic subunits PsbU and PsbV, was placed



**Figure 2** | Location of all the different subunits of the  $C_2S_2M_2$  particle. **a**, Overviews with the location of the core region (purple), the S-LHCII trimer, the M-LHCII trimer, CP29, CP26 and CP24 indicated. The top view shows PSII from the luminal side, normal to the membrane plane; the bottom view is along the membrane plane. **b**, Assigned subunits of the core, individually indicated in the figure. **c**, Close-up of the fit of PsbW in our structure. The lumen and stroma are indicated.

into the density map using a rigid body fit. Based on this initial fit, the homology models for the *Arabidopsis* subunits were then placed into the density map and refined as individual rigid bodies to adjust their relative positions and orientations. Each fit was examined individually, to determine whether the corresponding subunit was indeed present in the cryo-EM density map. The local fit and geometry of the core subunits were then optimized by model building and real space refinement at 4.5 Å resolution, using tight secondary structure restraints, side chain rotamer restraints and restraints to maintain a proper geometry and binding mode of the chlorophylls. Using a similar approach, separate 3D structures were obtained of the S trimers with subunits CP26 and CP29, and of the M trimers with subunits CP24, improving local fit and geometry by real space restrained refinement at 5.5 Å and 6.5 Å, respectively. The final PSII model was obtained by assembly of the separately built and refined regions (Fig. 2).

In the cryo-EM density map all the transmembrane helices of PsbA, PsbB, PsbC and PsbD are clearly visible and well defined. Whereas the helices are clearly visible, the termini are more difficult to assign. The region around these subunits is the best resolved part, and the porphyrin ring of the chlorophylls can be fitted accurately in the right plane.

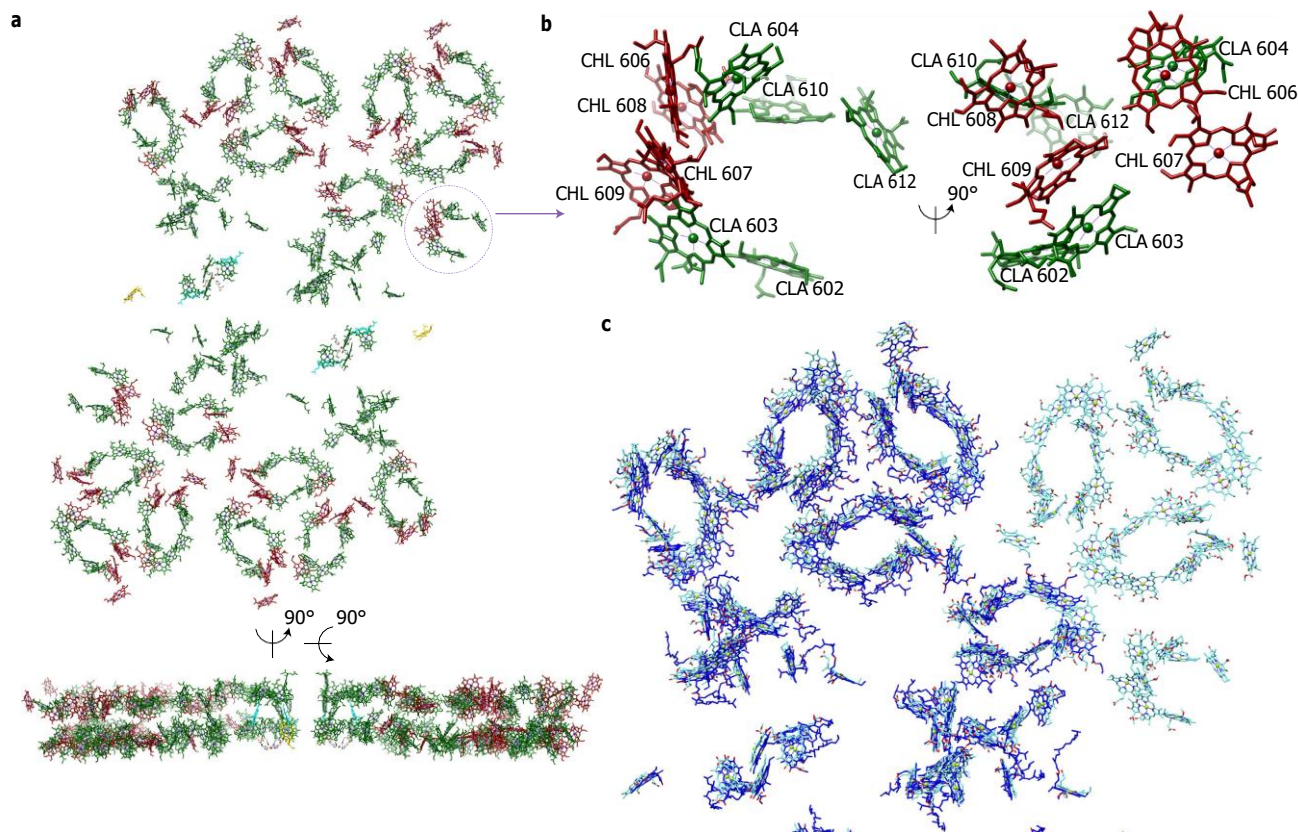
The small intrinsic subunits were analysed one by one. In the centre of the core the helices are well defined and the subunits PsbL, PsbM and PsbTc were easy to place. Towards the outside of the core it was more difficult, but overall most subunits could be assigned. The location of PsbE and PsbF, which coordinate the cytochrome *b*-559, could be confirmed by the clear presence of a density for the haem. The subunits PsbH, PsbI, PsbK, PsbX and PsbZ are clearly present as well. For the extrinsic subunits the density for PsbO has been assigned. The  $\beta$ -barrel is well organized, but for the loops not all densities are resolved. In the density map at each location of PsbO a single copy is present (Fig. 2a,b), as in the recent model of the  $C_2S_2$  particle from spinach<sup>8</sup>.

The other three extrinsic subunits, PsbQ, PsbP and PsbR, are absent in our map. The location of PsbP and PsbQ has been recently

proposed for the spinach PSII- $C_2S_2$  supercomplex<sup>8</sup>, but the location of PsbR remains still unknown. One unassigned density, in between PsbA, PsbC and the S trimer, is likely to be the subunit PsbW (Fig. 2c)<sup>8</sup>. To further confirm this assignment, the known amino acid sequence has been fitted in by taking into account the expected structure<sup>22</sup>. The full transmembrane part and a large part of the amino (N) terminus is modelled, starting from amino acid 9. The model indicates that PsbW has interactions with PsbA, PsbC and PsbI, and via its N terminus with PsbO. Possibly there is an interaction with a cofactor of the S trimer, because an unassigned density is present between PsbW and the S trimer subunit, above Chl *a*612. At the current resolution it is not possible to identify this density. Based on the size of the extra density we do not expect a chlorophyll at that location.

Two small membrane subunits, PsbJ and PsbY, could not be detected in the 3D map. No unassigned density was present in the electron microscopy density map at the position where PsbJ was expected according to the cyanobacterial structure. It should be noted that in the recent spinach PSII structure<sup>8</sup>, PsbJ has been assigned to a density between PsbK and PsbE. However this density is very weak, suggesting a sub-stoichiometric amount of PsbJ. To determine the location of PsbY, the cyanobacterial core structure<sup>5</sup> was used as a reference (pdb: 4UB6). At the putative location of PsbY (next to PsbK and PsbZ) a weak density is present. However, considering that the second transmembrane helix predicted in *Arabidopsis* (in cyanobacteria PsbY has only one helix) is not visible and that the resolution in the map at the possible location of PsbY is low, PsbY was not placed in the  $C_2S_2M_2$  model.

The structure gives insight into the location of the many pigments and co-factors, in particular for the four large subunits of the reaction centre that contain all the chlorophylls of the core. All 35 chlorophylls, the two pheophytins and the haem, which are present in cyanobacteria, were identified in the plant PSII core (Fig. 3). The chlorin rings of both the chlorophylls and the



**Figure 3 | Chlorophylls of the supercomplex.** **a**, Overview of all chlorophylls in the supercomplex shown from the luminal side and along the membrane plane. Chlorophyll *a* is depicted in green, chlorophyll *b* in dark red, pheophytins in cyan and haem in yellow. **b**, Close-up of CP24; a view from the lumen (left) and along the membrane plane (right). **c**, Overlay of the chlorophylls of the *Arabidopsis* C<sub>2</sub>S<sub>2</sub>M<sub>2</sub> particle (light blue) and the spinach C<sub>2</sub>S<sub>2</sub> particle (dark blue)<sup>8</sup>.

pheophytins are easy to locate, but for the phytol tails there is no clear density present and therefore they were removed from the model during the refinement. Owing to the missing phytol tails, it is not possible to determine the exact orientation of the chlorophylls, but there are no differences in distances between the chlorophylls in plants and cyanobacteria. This indicates that these chlorophylls are highly conserved between distant species.

All chlorophylls in the core could be localized, but other densities smaller than a chlorophyll remain unassigned. To have an idea where other pigments and lipids are located, pigments and lipids of cyanobacteria were overlaid on top of the unassigned densities.

Based on this analysis, it is possible to locate eight  $\beta$ -carotenes. In the cyanobacterial structure there are more than 20 lipids present<sup>4</sup> and in spinach there are 16 lipids, 10 carotenoids and 1 plastoquinone<sup>8</sup>; in our map the partial density of 10 of them could be assigned (data not shown). None of these pigments or lipids was included in the refined 3D model of PSII, as only partial densities are present in the electron microscopy map and the resolution is a limiting factor as well (Supplementary Fig. 5). The presence of a number of pigments or lipids could be confirmed, but the presence of the manganese cluster could not (Supplementary Fig. 6). All the amino acids which coordinate manganese, calcium or oxygen atoms are conserved in *Arabidopsis*. But the carboxy (C) terminus of PsbA (amino acids 335–344), important for the binding of the manganese cluster, appears to be disordered, as there is no density present in that region.

In this structure of the C<sub>2</sub>S<sub>2</sub>M<sub>2</sub> particle the location and orientation of the antenna proteins could be unambiguously determined although the overall resolution of the antenna region is not as high as in the core and among the different antenna

proteins the resolution is not consistent as well. The S trimer, CP26 and CP29 are the best-resolved components compared with the M trimer and CP24 (Supplementary Fig. 4). Therefore these regions were analysed separately. Chlorophylls could be detected in all these proteins, but no distinction could be made between Chl *a* and Chl *b*. The phytol tails of the chlorophylls were not used for the model as there is no density present. The carotenoids could not be detected at all. There are some unassigned densities in the S trimer, and it is possible that one pigment molecule is located there. But based on the current resolution it is impossible to determine if this is an extra pigment molecule or just a tail from one of the assigned chlorophylls.

Crystal structures of both the LHCII trimer and CP29 are available<sup>15,16</sup> and these were placed at the location of the S trimer and CP29. The trimer was analysed as a homotrimer with only Lhcb1 because of limited resolution of the S trimer, which does not allow a determination of the side chains of the amino acids. The individual amino acids that differ between pea and *Arabidopsis* were replaced; the same was done for the amino acids of CP29 that differ between spinach and *Arabidopsis*. Although the side chains of the amino acids are not resolved in the density, a rigid body fit followed by a real space refinement determined the precise orientation and location of both the S trimer and CP29. The root mean squared deviations (r.m.s.d.) between the backbone C $\alpha$  atoms of PSII-CP29 and those of the 3PL9 (ref. 16) and 2BHW (ref. 15) structures are 1.1 Å and 1.5 Å, respectively. Based on the location of the helices the structures could be placed rather exactly (Supplementary Fig. 7). Once the helices were placed, the location of the chlorophylls could be determined. In both structures all the chlorophyll binding sites, based on the

Table 1 | Detected chlorophylls of CP26 and CP24 and the chlorophyll binding site.

CP26*	CP24*	Coordinating residues (CP26/C24)
<i>b</i> 601 <sup>†</sup>		–
<i>a</i> 602	<i>a</i> 602	GLU114/GLU102
<i>a</i> 603	<i>a</i> 603	HIS117/HIS105
<i>a</i> 604	<i>a</i> 604	–/–
<i>b</i> 606	<i>b</i> 606	GLU178/GLN149
<i>b</i> 607	<i>b</i> 607	–/–
<i>b</i> 608	<i>b</i> 608	–/–
<i>a</i> 609	<i>b</i> 609	GLU186/GLU157
<i>a</i> 610	<i>a</i> 610	GLU225/GLU228
<i>a</i> 611		–/–
<i>a</i> 612	<i>a</i> 612	ASN228/HIS231
<i>a</i> 613		GLN242/–
<i>a</i> 614		HIS257/–

\*Chlorophyll types are based on the chlorophylls of CP29 (ref. 16) and biochemical data<sup>27</sup>. <sup>†</sup>Chlorophyll type is based on the trimer structure<sup>14</sup>.

backbone, are conserved. For the S trimer all 14 chlorophylls per monomer are present. For CP29, 13 chlorophylls were located, including Chl *a*601 and Chl *a*614, of which the last one is missing in the spinach CP29 cryo-EM map<sup>8</sup>. Chl *a*615 from the crystal structure<sup>16</sup> could not be located, as well as Chl *a*616 from spinach PSII<sup>8</sup>. The identity of chlorophyll 601 is based on the CP29 crystal structure and the identity of Chl 615, which is a Chl *a* assigned to a new site near to Chl 601 site in the N-terminal region. However, the N terminus in the CP29 crystal was severely affected during CP29 purification and crystallization, and Chl *a*615 was very likely to be a displaced Chl *a*601.

For CP26 a homology model was built, as there is no crystal structure available for this subunit. Once the exact location and orientation was determined, the chlorophylls could be located (Fig. 3). As the resolution does not allow differentiation between Chl *a* and Chl *b*, the same type of chlorophyll was assigned to every position as in CP29 or the LHCII trimer, although some adjustments were made based on biochemical data<sup>18,23,24</sup>. In this map 13 chlorophylls are detected. The chlorophyll binding sites are listed in Table 1. Based on the trimer structures, CP26 has a Chl *b*601. For all other locations the chlorophyll binding sites from the backbone are conserved. Just like CP29 there is no Chl 605 present. No other pigments could be detected in CP26.

The chlorophylls of the *Arabidopsis* and spinach structures were overlaid (Fig. 3c). When comparing these two plant structures there are hardly any differences. In both structures 35 chlorophylls in the core region are detected as well as all 14 chlorophylls of a single trimer subunit. Before these structures were solved it was assumed that there were 13 chlorophylls in CP29. The crystal structure contained a chlorophyll at location 615 which was unknown. In none of the plant structures a density was found for this chlorophyll but a density is present for a chlorophyll 601. The major difference between the two plant structures are the chlorophylls at location 614 and 616. In our structure we have found a chlorophyll at location 614 but none at location 616. Also there is no density that indicates that the presence of a chlorophyll at location 616. Conversely, in the spinach structure chlorophyll 614 is lost during purification but there is an extra chlorophyll present at location 616. This chlorophyll needs to be confirmed with biochemical data. For CP26 the same 13 chlorophylls have been found in both plant structures. This number of chlorophylls was expected based on the biochemical data<sup>18</sup>.

The density at the region for both the M trimer and CP24 is less ordered than in the other map regions (Supplementary Fig. 4). The M trimer was analysed as a homotrimer with only Lhcb1. A rigid

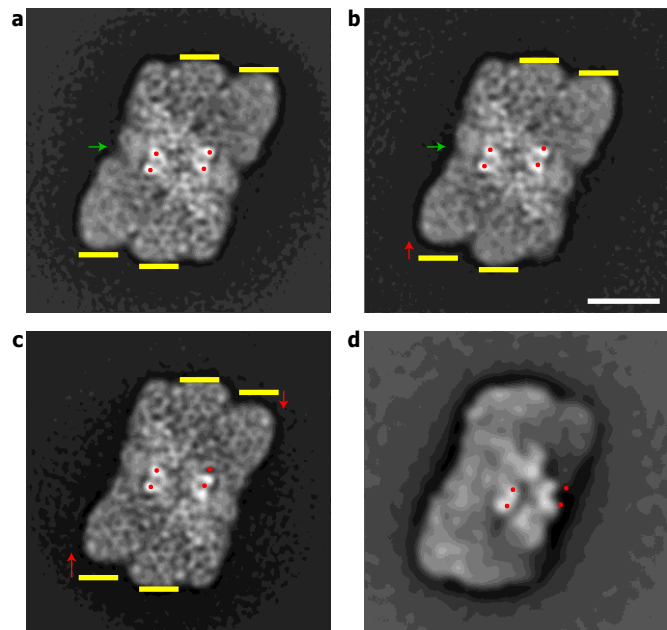


Figure 4 | Flexibilities in the  $C_2S_2M_2$  supercomplex from *Arabidopsis* investigated by negative stain single-particle electron microscopy. a, The sum of 28,048 particles of the standard  $C_2S_2M_2$  supercomplex. b, The sum of 22,626 particles of the supercomplex in which one M trimer is moved upward. Green arrows indicate a variable gap between CP24 and the core. c, The sum of 7,629 particles of the supercomplex in which both M trimers and the two halves of the core complex have been shifted, as can be seen from the positions of the central bright densities (red dots). d, The sum of 1,000 particles at a tilted view, in which the stain-excluding extrinsic subunits and the lumenal-protruding domains of CP47 and CP43 are apparent. Scale bar, 100 Å.

body fit could be applied for both CP24 and the M trimer. In the M trimer all the densities for the chlorophylls are present in all three individual monomers. For CP24 the individual chlorophylls had to be determined. The assignment of Chl *a* or Chl *b* was based on biochemical data<sup>24</sup> and the crystal structure of the homologue monomeric CP29 antenna<sup>16</sup>. For CP24, nine chlorophylls were detected (Table 1 and Fig. 3b). The chlorophyll binding sites of the backbone are conserved for the detected chlorophylls<sup>24</sup>. From biochemical data and sequence alignment, Chl 601, 613, 614 are likely absent in CP24<sup>24</sup>, and indeed these chlorophylls were not found in our electron microscopy map. In contrast, Chl 611 should be present<sup>24</sup>, but no density was detected for this pigment.

The flexibility within the peripheral antenna, apparently leading to a lower overall resolution in our map and in particular the flexibility of the M trimer/CP24 region is an interesting topic on its own. In addition to the cryo-electron microscopy data, we have analysed a set of 400,000 negatively stained *Arabidopsis* supercomplex projections (Fig. 4). Flexibility leads to a number of different, but specific conformations. It is found that the M trimer can have two clearly different positions, in contrast to the S trimer. When one of the two M trimers is shifted, there is a displacement of CP24 (Fig. 4b), leading to a smaller gap with the core (green arrow). This could explain a lower resolution of CP24 and the M trimer, than the PSII core complex and the S trimer/CP29 domain in the cryo-structure. In smaller numbers of particles, both M trimers are displaced (Fig. 4c). This leads to lateral movement of the two halves of the dimeric core complex, as indicated by the positions of red dots on the highest densities of the core part. These densities also change position on tilting (Fig. 4d), but not vertically as in Fig. 4b,c. The nature of the

M trimer flexibility is not understood since it is observed with a similar abundance in both dark-adapted plants and plants with light-induced non-photochemical quenching.

## Discussion

Over the years protein crystallization attempts have been made to obtain a high-resolution 3D structure for plant PSII, but the high flexibility and heterogeneity of PSII supercomplexes has prevented success. Single-particle cryo-EM was applied on a homogeneous sample of  $C_2S_2M_2$  particles from a *npq4* mutant, lacking PsbS. Use of the mutant is justified because previous electron microscopy analyses on PSII particles and membranes from wild-type and *npq4* plants did not reveal any structural difference in isolated PSII<sup>20</sup>. Moreover, it was shown that maximum quantum yield of PSII, quantum yield of oxygen evolution, maximal oxygen production rate and pigment content are identical in wild-type and *npq4* plants<sup>21,25–27</sup>. We managed to determine a 3D structure for the  $C_2S_2M_2$  supercomplex to 5.3 Å resolution. Based on the structure we are able to describe the location of almost all expected intrinsic membrane embedded subunits of the core. In this electron microscopy map the location of the 35 chlorophylls of the core based on the cyanobacterial structure<sup>4</sup> were confirmed. This result matches the results of the spinach structure as well<sup>8</sup>, indicating that these chlorophylls are highly conserved between different organisms. Of the extrinsic subunits only PsbO is present. PsbQ, PsbP and PsbR are absent in our map: the location of PsbP and PsbQ has been recently proposed<sup>8</sup>, but the location of PsbR remains unknown. The absence of three of the extrinsic subunits and some other small intrinsic subunits, as discussed below, is a consequence of the way to keep the particles in a monodisperse state for cryo-EM. Hence the structure does not provide insight into the oxygen evolution, which has been described in detail for the system from cyanobacteria<sup>4,5</sup>. The location of the chlorophylls, however, provides insight into the antenna organization and the energy flow in the membrane compartment of PSII.

In addition, the position of all peripheral antenna proteins in the largest purifiable PSII supercomplex could be determined, including CP24 and CP26. Since no crystal structure is available for these two subunits, electron microscopy densities allowed significant information to be obtained about protein structure and pigment organization to build a 3D model of these proteins.

All densities could be assigned to the small subunits, but PsbJ and PsbY are missing from the map. It is unlikely that one of these subunits is present. PsbJ is a very hydrophobic chloroplast-encoded protein and operates in a gene cluster with *psbE*, *psbF* and *psbL*. In the structures of *Thermosynechococcus vulcanus* PsbJ is placed between PsbE/PsbF and PsbK<sup>4,5</sup> and, as PsbJ functions in cyanobacteria and higher plants in a similar way<sup>28</sup>, there is no indication that it is relocated to a position elsewhere. PsbW was assigned next to PsbA, PsbC, PsbI and S-LHCII based on a good match with the expected structure (Fig. 2d). This location of PsbW, between the core and the antenna proteins, is in agreement with the map of spinach PSII<sup>8</sup> and previous results because in the absence of PsbW no supercomplexes are formed<sup>29</sup>.

The aim of this study was to perform a reconstruction on a purified supercomplex particle with a complete set of subunits, including the four extrinsic subunits of the OEC (PsbO, PsbP, PsbQ and PsbR). The sucrose gradient centrifugation and concentration step were carried out at pH 7.5, to prevent aggregation. As indicated elsewhere, this relatively high pH value may destabilize the OEC<sup>9</sup>, which appears to be the case for our digitonin preparation<sup>30</sup>. This could explain why both PsbP and PsbQ are missing from the structure. PsbR is absent as well, although it is more difficult to wash off this subunit<sup>30</sup>. PsbR needs PsbJ for its assembly into the PSII core<sup>31</sup>, and the structure clearly shows the absence of PsbJ. Moreover, the absence of PsbR further affects its neighbours PsbP

and PsbQ<sup>10</sup>. The fact that most extrinsic subunits are missing from our structure could probably have destabilized the domain around the manganese cluster, explaining why the cluster is not visible in the 3D map.

It should be noted that biochemical characterization of digitonin-purified PSII particles from *Arabidopsis* suggests that PsbR (and possibly also PsbP and PsbQ) is present in substoichiometric amount<sup>30</sup>, not only in the purified PSII supercomplex, but also in PSII in leaves<sup>30</sup>. PsbR is also lacking in the recent cryo-EM structure of the smaller  $C_2S_2$  PSII complex of spinach<sup>8</sup>, possibly for the same reason as for *Arabidopsis* particles (substoichiometric amount of PsbR). Thus, further investigation is needed to understand the stoichiometry, function and location of PsbR. Conversely, in the  $\alpha$ -dodecyl maltoside-purified PSII particles of spinach, PsbP and PsbQ are visible<sup>8</sup>. However also in this case PsbP and PsbQ are present in substoichiometric amount. Indeed their visualization requires decreasing the level of electron microscopy density threshold to a very low value. This usually significantly increases the noise signal, but the fact that these peripheral subunits do not bind detergent allows their visualization with little noise compared with the rest of the PSII. It is also possible that PsbP and PsbQ are more strongly bound to the PSII core in spinach than in *Arabidopsis*.

The PsbS protein has the key function of activating the fast energy dissipation at the level of PSII in a still unknown way. Thus, localization of PsbS in the supercomplex is necessary to understand its mechanism of action. In our work, we cannot provide data about PsbS localization since, as discussed above, we chose to use PSII from *npq4* plants lacking PsbS. So far no data show a role of PsbS in electron transfer mechanisms or its interaction with OEC subunits, but rather recent biochemical evidences support the fact that PsbS could interact with the external antennas<sup>32,33</sup>, as proposed for some time. Moreover, previous work by us and other groups did not show any difference in functioning and structure between wild-type and *npq4* PSII, as long as plants are grown in non-stressed conditions. This means that PsbS does not have an evident structural role in PSII, at least under non stressing conditions. Intriguingly, the localization of PsbS remains elusive despite the fact that this is an objective of several groups for over 15 years. All assays to localize PsbS in the supercomplex by electron microscopy have not had success<sup>13,34</sup>, indicating that PsbS is not or just very weakly bound to PSII. It should also be noted that in the recent spinach cryo-EM structure<sup>8</sup>, PsbS was present in the fraction containing the  $C_2S_2$  particles (although the stoichiometry was not calculated), but it was not found in the density map of the complex, further indicating that PsbS is not, or very weakly, bound to PSII. Another possibility might be that PsbS is bound at several different positions around PSII.

Not all protein subunits were identified, but all chlorophylls present in cyanobacteria are identified in the plant PSII map and all chlorophyll positions of cyanobacterial PSII are highly conserved in plants. These are the same chlorophylls as in the spinach structure, and all the chlorophylls have the same orientation in the two plant structures. In our density map there are unassigned densities in the core. According to their size it is likely to be that these unassigned densities are carotenoids and lipids, rather than chlorophylls although it has been suggested earlier that PSII core complexes in plants would have extra chlorophylls compared with cyanobacteria PSII<sup>18</sup>.

## Orientation and structure of antenna proteins

The location and orientation of the antenna complexes could be determined unambiguously from our map. The positions of most of the antenna complexes are similar to those proposed earlier in the 2D negative stain map<sup>20</sup>, but the current data are much more accurate. Only a 3D map with a resolution sufficient to at least resolve individual  $\alpha$ -helices allows modelling the correct location

and orientation of LHCII and the minor antenna subunits. Moreover, membrane proteins cannot be determined accurately from negative stain data because the stain does not penetrate the hydrophobic membrane. A comparison of the two models shows that the S- and M trimers are rotated by approximately 5° (Supplementary Fig. 8). The exact orientation of CP26 and CP29 is slightly different from the previous model, but the orientation of CP24 is significantly different. We can now also compare the positions of CP26 and CP29 with those of a high-resolution structure of the smaller C<sub>2</sub>S<sub>2</sub> particle from spinach<sup>8</sup>. Although at low resolution the spinach C<sub>2</sub>S<sub>2</sub> particle often looks closer to a rectangle than the *Arabidopsis* C<sub>2</sub>S<sub>2</sub> particle<sup>20</sup>, it now appears that in both supercomplexes the CP26 subunit and the S trimer are in a very similar position.

Compared to the S trimer, the M trimer is not well resolved, although all chlorophylls could be detected. There is good evidence from the negative stain maps (Fig. 4) that the position of the M trimer/CP24 is variable, leading to a lower local resolution. In a small number of supercomplexes there is also an additional shift of the two halves of the core. Therefore the exact orientation and location in the model for the supercomplex may change once this region of the map becomes resolved at a higher resolution. In the current map it is possibly a mixture of two or more conformations, but to separate them a larger dataset is necessary. LHCII-M specifically contains one Lhcb3 monomer, an LHCII isoform evolved in land plants in parallel with CP24 (ref. 35). In some plant clade, Lhcb3 recently disappeared together with CP24 (ref. 36), indicating a relationship between these two Lhcb complexes. The resolution of the M-LHCII is not good enough to distinguish the specific Lhcb3 features (like a shorter N terminus and the specific amino acid substitutions with respect to Lhcb1). However, the relative orientation of M-LHCII and CP24 obtained from the PSII model indicates that a contact of CP24 with M-LHCII involves the region between helix B and helix E of the LHCII monomer facing CP24. Interestingly, this is the most variable region between Lhcb1 and Lhcb3, suggesting that the monomer in contact with CP24 is indeed Lhcb3 and its specific residues are important for the interaction.

CP29 is slightly tilted relative to the plane of the membrane, which may have implications for the expected energy flow within PSII. 13 chlorophylls were detected for CP29 in our density map, and the same number was also present in the crystal structure. However, there are important differences. First, Chl 615 is missing in the electron microscopy structures of *Arabidopsis* and spinach PSII, and there is no unassigned density near the expected location of Chl 615 that could account for a chlorophyll molecule. Our results support the conclusion that the position of Chl 615 in the crystal structure is an artefact, because the N terminus of CP29 is not intact as it is in the cryo-EM structures. Second, at the position corresponding to Chl 601 in LHCII there is density for an extra chlorophyll. The resolution in the *Arabidopsis* and spinach maps is not high enough for identification of the Chl species. We assigned Chl 601 as Chl *a*, based on the identify of Chl 615 in the spinach CP29 crystal structure, since it is very likely to be that Chl 615 is Chl 601 displaced during CP29 purification and crystallization. This is the same assignment as in CP29 in the spinach PSII structure<sup>8</sup>. In LHCII this chlorophyll was however assigned as Chl *b*<sup>14,15</sup>. Higher resolution is necessary to solve this question.

There are two significant differences between CP29 in *Arabidopsis* and spinach PSII: Chl614 is present in CP29 of *Arabidopsis*, but absent in the spinach CP29. For Chl 616, it is the opposite situation, because we have not found any extra density that could indicate the presence of Chl 616. Since the Chl614 presence is proven from biochemical experiments<sup>37</sup> and the crystal structure<sup>14</sup>, it is likely to be that this chlorophyll is lost from spinach PSII during preparation.

In our map 13 chlorophylls are detected in CP26. The same number of chlorophylls was predicted earlier<sup>18</sup>. The known crystal structures of both CP29 and a single subunit of the LHCII trimer where overlaid on the location of CP26 to identify the chlorophylls. Based on these overlays we concluded that a chlorophyll is present at position 601. Similar to CP29, there is no chlorophyll present at position 605. All other chlorophyll binding sites are conserved. When comparing this structure with the one from spinach<sup>8</sup>, there is no significant difference in chlorophyll location or orientation between the two structures.

The location and the most likely orientation of the CP24 protein were determined. Of the 10–11 predicted chlorophylls<sup>18,24</sup>, 9 were detected at conserved binding sites. Surprisingly, Chl 611 was not detected. This chlorophyll, which is not coordinated by a protein residue, is predicted from biochemical data<sup>24</sup> and conserved in other Lhc complexes. Therefore, it is possible that a density was not found for the low resolution of this part of the complex and/or its loss during PSII purification. Chl 614 is absent, in accordance with the lack of the binding site due to a shorter C terminus. A residue for Chl 613 binding could be present in the protein sequence<sup>24</sup>, but there is no density for this chlorophyll in the 3D map. This result supports biochemical results proposing its absence<sup>24</sup>. Concerning Chl 601, no biochemical or structural data allow concluding about its presence or not. The density of Chl *a*603 is weak, but based on its conserved binding site in all Lhcs, it was placed in the map. To identify possible missing chlorophylls in this minor antenna protein a better resolution is required.

### Excitation energy transfer

Detailed investigation and simulation of energy transfer kinetics in large pigment systems as plant PSII, which contains >300 chlorophylls, are extremely complex and require precise orientations of the chlorophylls and approaches beyond the scope of this article. Recent examples focus on plant PSII kinetics using an advanced modelling approach<sup>38,39</sup> based on a previous pseudo-atomic model of PSII<sup>20</sup> and exciton systems of individual subunits. However, the improved PSII structural model here presented allows providing qualitative information and preferential pathways for energy transfer between Lhc antennas and from Lhc towards the core (Fig. 5 and Table 2) based on the Chl species and distances between the nearest chlorophylls in adjacent complexes. This is the same approach as used for pea PSI<sup>40</sup> and spinach PSII complexes<sup>8</sup>. The almost identical organization of the *Arabidopsis* and spinach PSII core complex, as well as the further similarity between plant and cyanobacterial PSII, make it very likely that energy transfer inside the core complexes is similar between plant and cyanobacteria PSII. This topic has been investigated previously<sup>41</sup> and therefore we limit our discussion to the Lhc antenna system.

Efficient connection and energy transfer between S- and M-LHCII trimers is more likely to be provided between three couples of Chl *a*: Chl *a*611<sub>(S3)</sub>/Chl *a*610<sub>(M2)</sub>, Chl *a*612<sub>(S3)</sub>/Chl *a*612<sub>(M2)</sub> and Chl *a*614<sub>(S3)</sub>/Chl *a*604<sub>(M2)</sub> (Fig. 5 and Table 2). It should be noted that Chls 611 and 612, located near the stromal side of LHCII, have the lowest energy states in all Lhcb complexes and are the most populated at equilibrium. From the S trimer, excitation energy can directly and efficiently flow to CP43 from Chl *a*611/*a*612<sub>(S1)</sub> to Chl *a*506<sub>(CP43)</sub> and from Chl *a*614<sub>(S1)</sub> to Chl *a*501<sub>(CP43)</sub>. From M trimer to CP29, energy migration involves the lowest energy chlorophylls of this two complexes, in particular Chl *a*611<sub>(M1)</sub> and Chl *a*611<sub>(CP29)</sub>. It should be noted that the shortest distance between chlorophylls of the M trimer and CP29 is between Chl *a*614<sub>(M1)</sub> and Chl *b*614<sub>(CP29)</sub> on the luminal side of LHCII, but likely this pathway is less populated since it involves in CP29 a Chl *b*, which has a higher energy level than Chl *a*. However it should be noted that a recent hypothesis suggests that conformational changes in the helix 5/Chl 614 domain may form an energization quenching (qE) site in Lhc antennas<sup>42</sup>.





Table 2 | Distance between the Mg–Mg of the chlorophylls involved in the energy transfer.

(A)		(B)					
Subunit and Chl	Subunit and Chl	Mg–Mg distance (Å)	Location	Subunit and Chl	Subunit and Chl	Mg–Mg distance (Å)	Location
M2-LHCII	S3-LHCII						
Chl a610	Chl a611	19.8	Stroma				
Chl a612	Chl a612	23.1	Stroma				
Chl a604	Chl a614	20.9	Lumen				
M1-LHCII	CP29			M1-LHCII	CP29		
Chl a611	Chl a611	17.1	Stroma	Chl a614	Chl b614	13.8	Lumen
CP29	CP47						
Chl a603	Chl a610	18.2	Stroma				
Chl a609	Chl a616	19.8	Stroma				
Chl a613	Chl a602	20.5	Lumen				
S1-LHCII	CP43			S1-LHCII	CP29		
Chl a611	Chl a506	17.2	Stroma	Chl b605	Chl b606	19.1	Lumen
Chl a612	Chl a506	20.2	Stroma	Chl b605	Chl a604	17.3	Lumen
Chl a614	Chl a501	24.5	Lumen				
S2-LHCII	CP26			S2-LHCII	CP26		
Chl a604	Chl a604	24.4	Lumen	Chl b605	Chl a604	19.5	Lumen
Chl a610	Chl a610	25.0	Stroma	Chl b608	Chl a610	20.5	Stroma
				Chl b608	Chl a612	20.7	Stroma
CP26	CP43			CP26	CP43		
Chl a611	Chl a513	15.0	Stroma	Chl b601	Chl a513	11.1	Stroma
Chl a611	Chl a512	18.2	Stroma				
Chl a614	Chl a503	15.9	Lumen				
CP24	M1-LHCII			CP24	M1-LHCII		
Chl a610	Chl a612	24.2	Stroma	Chl b608	Chl a612	20.8	Stroma
Chl a604	Chl a604	25.9	Lumen				
				CP24	CP29		
				Chl b606	Chl b614	19.0	Lumen
				Chl b608	Chl a601	18.2	Stroma
				Chl b609	Chl a601	20.1	Stroma

(A) Energy transfer pathways involving couples of Chl *a*. (B) Energy transfer pathways involving one or two Chl *b*. In each case, the nearest chlorophyll of two adjacent subunits is indicated together with the Mg–Mg distance and the luminal or stromal location. The names of the single LHCII monomers are given as indicated in Fig. 5.

CP43, while from the M trimer it needs to pass through CP29 to reach the core on CP47 or through S-LHCII to reach CP43. Despite the fact that PSII is considered rather isoenergetic and thus energy can move randomly between subunits, the fact that monomeric CP26 and CP29 seem well connected with the core and less well connected with S trimer could provide some directionality of the energy migration towards the core and limit the equilibration on the antenna system.

Lastly, it is worth noting that the distances between chlorophylls in PSII antennas are larger than in PSI<sup>3,40</sup>, which indeed, despite the presence of low energy chlorophylls (red forms) in the antenna system, has a very fast energy transfer from LHCI to the core, faster than the one proposed between Lhcb antennas and the PSII core<sup>43</sup>. It is also worth noting that most of the energy connections are on the stromal side of the Lhc complexes. This could facilitate *in vivo* energy exchange between PSII located in opposite highly stacked grana membranes, further increasing the connectivity of PSII.

## Methods

**Protein purification.** PSII-enriched membranes (BBY)<sup>44</sup> were prepared from *A. thaliana npq4* plants, following the same protocol<sup>20</sup>. PSII-C<sub>2</sub>S<sub>2</sub>M<sub>2</sub> particles were prepared as by Crepin and coworkers<sup>30</sup>, where further biochemical characterization is provided. In short, aliquots of 200 µg (in chlorophylls) of BBY membranes were washed with 10 mM Hepes KOH pH 7.5, resuspended at a concentration of 1 mg ml<sup>-1</sup> in 10 mM Hepes KOH pH 7.5 and solubilized by adding an equal volume of 1% digitonin and 0.4% α-dodecylmaltoside to a final chlorophyll concentration of 0.5 mg ml<sup>-1</sup> and a final detergent concentration of 0.5% digitonin and 0.2% α-dodecylmaltoside. The samples were incubated on ice in the dark for 30 min and then centrifuged at 12,000g for 10 min to eliminate unsolubilized material. PSII supercomplexes were purified by sucrose gradient centrifugation with a sucrose gradient containing 0.65 M sucrose, 10 mM Hepes KOH pH 7.5 and 0.01% digitonin. Gradients were formed by freezing and thawing. Ultracentrifugation was performed in a Beckmann SW41 rotor, at 41,000 r.p.m. for 17 h.

PSII supercomplexes for electron microscopy of negatively stained specimens were isolated using clear-native polyacrylamide gel electrophoresis<sup>45</sup> with the following modifications. Thylakoid membranes with 10 µg of chlorophylls were solubilized with α-dodecylmaltoside using a detergent/chlorophyll mass ratio of 40 and supplemented with sample buffer (50 mM HEPES pH 7.2, 400 mM sucrose, 5 mM MgCl<sub>2</sub>, 15 mM NaCl, 10% glycerol) to the final volume of 30 µl.

**Sample preparation and electron microscopy.** A volume of 6 ml of sucrose PSII solution at a concentration of 0.033 µg ml<sup>-1</sup> chlorophyll was concentrated about 100× to a final concentration of 3.5 mg ml<sup>-1</sup> chlorophyll. The sample was placed on a Millipore Amicon filter (10 kDa cut-off) and centrifuged at 3,500g for 20 min per round. After every washing round the sample was diluted with 2 ml of buffer (10 mM Hepes, 0.01% digitonin). This was repeated six times, a final seventh run was used to concentrate the sample to a final volume of 60 µl. Aliquots of 3 µl of the concentrated PSII sample were applied onto glow-discharged holey carbon grids (Quantifoil R1.2/1.3). Grids were blotted for 8 s at 100% humidity, before being plunge-frozen in a FEI Vitrobot using liquid ethane. Grids were transferred to a G2-Polaris (FEI), operating at 300 kV, equipped with a Gatan energy filter and a Gatan 4000 SP 2 K CCD camera. Micrographs were recorded with a pixel size of 2.63 Å at the specimen level, with a defocus of 3 µm and with a dose of 25 e<sup>-</sup> Å<sup>-2</sup> (electrons per square Ångström) and 25,996 particles were manually selected with Xmipp 3.0<sup>46</sup> or RELION-1.3<sup>47</sup>.

For high-resolution electron microscopy, grids were transferred to a FEI Titan Krios (NeCEN, Leiden), operating at 300 kV, equipped with a Falcon 2 direct electron detector. Micrographs were collected at a pixel size of 1.105 Å at the specimen level and an exposure time of 1 s, corresponding with an electron dose of 28, 38 or 49 e<sup>-</sup> Å<sup>-2</sup> for the integrated image. The frames had a dose of 3, 4.2 and 5.4 e<sup>-</sup> Å<sup>-2</sup>, respectively. A total of 5,198 images, with a defocus range of 1.2–3.0 µm, were collected with EPU software (FEI) and 104,025 particles selected.

Single-particle electron microscopy of negatively stained specimens was performed on a Tecnai G2 20 Twin transmission electron microscope (FEI, Eindhoven, the Netherlands), operated at 200 kV. Images were recorded with an UltraScan 4000UHS CCD camera (Gatan, Pleasanton, CA, USA) with a pixel size of 2.24 Å at the specimen level after binning the images to 2048 × 2048 pixels.

**Data processing.** From the G2-Polaris dataset the best 13,570 particles were used for an initial 3D model, made with Ransac<sup>48</sup>, using C2 symmetry. During every 3D

classification and 3D refinement C2 symmetry was applied. This initial 3D model was further refined and used as a reference for the dataset of the Titan Krios. The contrast transfer function parameters of the G2-Polaris micrographs were determined with CTFFIND<sup>39</sup>. CTFFIND<sup>40</sup> was used on the summed images of the frames aligned by MOTIONCORR<sup>31</sup>, to correct for beam induced motion and drift. The dataset was processed with RELION-1.3, two initial 3D classifications were used to clean the datasets based on a three times binned pixel (3.32 Å), to select the best 23,434 particles combined for a 3D refinement. Many of the rejected particles lacked a LHCII trimer. As a reference a 40 Å low-pass filtered initial model was used. The 3D classification parameters were used for a 3D refinement at twice-binned pixel (2.21 Å) and on the original pixel (1.105 Å). With a twice-binned pixel size the resolution after 3D refinement was 7.0 Å. For the original pixel after particle polishing and refinement<sup>52</sup> the best particles gave a resolution of 5.3 Å. At the original pixel size a map with a resolution of 7.0 Å was low-pass filtered to 40 Å and used as a reference. All density maps were corrected with the modulation transfer function of the detector and sharpened by applying a negative B-factor (−60)<sup>53</sup>. All resolutions are based on the gold-standard Fourier shell correlation (FSC) = 0.143 criterion<sup>54</sup>. Local resolutions of the density maps were calculated in ResMap<sup>55</sup>. From the Tecnai G2 data set almost 400,000 particles were analysed using Xnipp<sup>56</sup> and Relion software<sup>47</sup>.

Modelling and bioinformatics tools. Initial fitting of the subunits in the cryo-EM map was performed by rigid body real space refinement in Chimera<sup>57</sup>, using as templates the high-resolution crystal structures of *Thermosynechococcus vulcanus* PSII (PDB code 3WU2)<sup>4</sup>, pea LHC-II (PDB code 2BHW<sup>15</sup>) for the S- and M trimers and spinach CP29 (PDB code 3PL9<sup>16</sup>) for CP29, CP26 and CP24. Local fitting and adjustment of the subunits in the cryo-EM maps was performed using the programs Coot<sup>58</sup> and phenix.real\_space\_refine<sup>59,60</sup>, starting from homology models constructed using the Phyre2 server<sup>61</sup>. Model refinement was carried out using symmetry and geometry restraints, including secondary structure, rotamer and Ramachandran plot restraints. Secondary structure assignments were based on analysis of the structures with DSSP<sup>62</sup>. Geometries of the chlorophylls and haems were strictly restrained towards those found in the high-resolution crystal structures, including correct coordination geometry of the metals with protein residues. A model of PsbW was obtained by fitting a polyaniline  $\alpha$ -helix into the empty density for a transmembrane helix. After one round of model building and refinement, side chains were replaced to match the actual amino acid sequence of PsbW. Owing to large differences in local resolution of the cryo-EM map, refinement of the PSII core, the S trimer with CP26/CP29 and the M trimer with CP24 was performed separately in excised parts of the cryo-EM map. Refinement of the PSII core was carried out at 4.5 Å resolution; the S trimer with CP26/CP29 and the M trimer with CP24 were refined at 5.5 Å and 6.5 Å resolution, respectively. The final model was created by reassembling the independently refined structural regions without further refinement of the complete structure. The overall clashscore of the complete PSII structure is 8, with 0.0% Ramachandran outliers and 0.1% side chain rotamer outliers, as calculated by Molprobity<sup>63</sup>, indicated that the geometry of the structure is acceptable. High-resolution images for publication were prepared with Chimera<sup>57</sup> and PyMOL (Molecular Graphics System, version 1.8 Schrödinger, LLC).

Data availability. The data that support the findings of this study are available from the corresponding author upon request. The Cryo-EM map and refined 3D structure of PSII were deposited at the Electron Microscopy Data Bank (EMDB) and Protein Data Bank (PDB) with accession codes [EMD-3491](#) and [5MDX](#), respectively.

Received 15 November 2016; accepted 24 April 2017;  
published 12 June 2017

## References

- Dekker, J. P. & Van Grondelle, R. Primary charge separation in photosystem II. *Photosynth. Res.* 63, 195–208 (2000).
- Nelson, N. & Yocum, C. F. Structure and function of photosystems I and II. *Ann. Rev. Plant Biol.* 57, 521–565 (2006).
- Mazor, Y., Borovikova, A. & Nelson, N. The structure of plant photosystem I super-complex at 2.8 Å resolution. *eLife* 4, 213 (2015).
- Ferreira, K. N., Iverson, T. M., Maghlaoui, K., Barber, J. & Iwata, S. Architecture of the photosynthetic oxygen-evolving center. *Science* 303, 1831–1838 (2004).
- Umena, Y., Kawakami, K., Shen, J. & Kamiya, N. Crystal structure of oxygen-evolving photosystem II at a resolution of 1.9 Å. *Nature* 473, 55–60 (2011).
- Suga, M. *et al.* Native structure of photosystem II at 1.95 Å resolution viewed by femtosecond X-ray pulses. *Nature* 517, 99–103 (2014).
- Ago, H. *et al.* Novel features of eukaryotic photosystem II revealed by its crystal structure analysis from a red alga. *J. Biol. Chem.* 291, 5676–5687 (2016).
- Wei, X. *et al.* Structure of spinach photosystem II–LHCII supercomplex at 3.2 Å resolution. *Nature* 534, 69–74 (2016).
- Bricker, T. M., Roose, J. L., Fagerlund, R. D., Frankel, L. K. & Eaton-Rye, J. J. The extrinsic proteins of photosystem II. *Biochim. Biophys. Acta* 1817, 121–142 (2012).
- Ifuku, K. Localization and functional characterization of the extrinsic subunits of photosystem II: an update. *Biosci. Biotechnol. Biochem.* 79, 1223–1231 (2015).
- Caffarri, S., Croce, R., Cattivelli, L. & Bassi, R. A look within LHCII: differential analysis of the Lhcb1–3 complexes building the major trimeric antenna complex of higher-plant photosynthesis<sup>1</sup>. *Biochemistry* 43, 9467–9476 (2004).
- Kouřil, R., Dekker, J. P. & Boekema, E. J. Supramolecular organization of photosystem II in green plants. *Biochim. Biophys. Acta* 1817, 2–12 (2012).
- Nosek, L., Semchonok, D., Boekema, E. J., Ilik, P. & Kouril, R. Structural variability of plant photosystem II megacomplexes in thylakoid membranes. *Plant J.* 89, 104–111 (2017).
- Liu, Z. F. *et al.* Crystal structure of spinach major light-harvesting complex at 2.72 Å resolution. *Nature* 428, 287–292 (2004).
- Standfuss, J., Terwisscha van Scheltinga, A. C., Lamborghini, M. & Kühlbrandt, W. Mechanisms of photoprotection and nonphotochemical quenching in pea light-harvesting complex at 2.5 Å resolution. *EMBO J.* 24, 919–928 (2005).
- Pan, X. *et al.* Structural insights into energy regulation of light-harvesting complex CP29 from spinach. *Nat. Struct. Mol. Biol.* 18, 309–315 (2011).
- Pan, X., Liu, Z., Li, M. & Chang, W. Architecture and function of plant light-harvesting complexes II. *Curr. Opin. Struct. Biol.* 23, 515–525 (2013).
- Caffarri, S., Tibiletti, T., Jennings, R. C. & Santabarbara, S. A. Comparison between plant photosystem I and photosystem II architecture and functioning. *Curr. Protein Pept. Sci.* 15, 296–331 (2014).
- Nogales, E. & Scheres, S. H. W. Cryo-EM: a unique tool for the visualization of macromolecular complexity. *Mol. Cell* 58, 677–689 (2015).
- Caffarri, S., Kouřil, R., Kereiche, S., Boekema, E. J. & Croce, R. Functional architecture of higher plant photosystem II supercomplexes. *EMBO J.* 28, 3052–3063 (2009).
- Li, X. P., Muller-Moule, P., Gilmore, A. M. & Niyogi, K. K. PsbS-dependent enhancement of feedback de-excitation protects photosystem II from photoinhibition. *Proc. Natl Acad. Sci. USA* 99, 15222–15227 (2002).
- Shi, L. X., Lorkovic, Z. J., Oelmüller, R. & Schröder, W. P. The low molecular mass PsbW protein is involved in the stabilization of the dimeric photosystem II complex in *Arabidopsis thaliana*. *J. Biol. Chem.* 275, 37945–37950 (2000).
- Caffarri, S., Passarini, F., Bassi, R. & Croce, R. A specific binding site for neoxanthin in the monomeric antenna proteins CP26 and CP29 of photosystem II. *FEBS Lett.* 581, 4704–4710 (2007).
- Passarini, F., Wientjes, E., Hienrwadel, R. & Croce, R. Molecular basis of light harvesting and photoprotection in CP24: unique features of the most recent antenna complex. *J. Biol. Chem.* 284, 29536–29546 (2009).
- Golan, T., Muller-Moule, P. & Niyogi, K. K. Photoprotection mutants of *Arabidopsis thaliana* acclimate to high light by increasing photosynthesis and specific antioxidants. *Plant Cell Environ.* 29, 879–887 (2006).
- Li, X. P. *et al.* A pigment-binding protein essential for regulation of photosynthetic light harvesting. *Nature* 403, 391–395 (2000).
- Roach, T. & Krieger-Liszskay, A. The role of the PsbS protein in the protection of photosystems I and II against high light in *Arabidopsis thaliana*. *Biochim. Biophys. Acta* 1817, 2158–2165 (2012).
- Shi, L., Hall, M., Funk, C. & Schröder, W. P. Photosystem II, a growing complex: updates on newly discovered components and low molecular mass proteins. *Biochim. Biophys. Acta* 1817, 13–25 (2012).
- Garcia-Cerdan, J. G. *et al.* The PsbW protein stabilizes the supramolecular organization of photosystem II in higher plants. *Plant J.* 65, 368–381 (2011).
- Crepin, A., Santabarbara, S. & Caffarri, S. Biochemical and spectroscopic characterization of highly stable photosystem II supercomplexes from *Arabidopsis*. *J. Biol. Chem.* 291, 19157–19171 (2014).
- Suorsa, M. *et al.* Psbr, a missing link in the assembly of the oxygen-evolving complex of plant photosystem II. *J. Biol. Chem.* 281, 145–150 (2006).
- Gerotto, C., Franchin, C., Arrigoni, G. & Morosinotto, T. *In vivo* identification of photosystem II light harvesting complexes interacting with Photosystem subunit S. *Plant Physiol* 168, 1747–1761 (2015).
- Correa-Galvis, V., Poschmann, G., Melzer, M., Stuehler, K. & Jahns, P. Psbs interactions involved in the activation of energy dissipation in *Arabidopsis*. *Nat. Plants* 2, 15225 (2016).
- Pagliano, C. *et al.* Proteomic characterization and three-dimensional electron microscopy study of PSII-LHCII supercomplexes from higher plants. *Biochim. Biophys. Acta* 1837, 1454–1462 (2014).
- Alboresi, A. *et al.* *In silico* and biochemical analysis of *Physcomitrella patens* photosynthetic antenna: identification of subunits which evolved upon land adaptation. *PLoS ONE* 3, e2033 (2008).
- Kouřil, R., Nosek, L., Bartos, J., Boekema, E. J. & Ilik, P. Evolutionary loss of light-harvesting proteins Lhcb6 and Lhcb3 in major land plant groups—Break-up of current dogma. *New Phytol.* 210, 808–814 (2016).
- Bassi, R., Croce, R., Cugini, D. & Sandona, D. Mutational analysis of a higher plant antenna protein provides identification of chromophores bound into multiple sites. *Proc. Natl Acad. Sci. USA* 96, 10056–10061 (1999).
- Kreisbeck, C. & Aspuru-Guzik, A. Efficiency of energy funneling in the photosystem II supercomplex of higher plants. *Chemical Sci.* 7, 4174–4183 (2016).
- Bennet, D. I. G., Amarnath, K. & Fleming, G. R. A structure-based model of energy transfer reveals the principles of light harvesting in photosystem II supercomplexes. *J. Am. Chem. Soc.* 135, 9164–9173 (2013).

40. Qin, X., Suga, M., Kuang, T. & Shen, J. R. Structural basis for energy transfer pathways in the plant PSI-LHCI supercomplex. *Science* 348, 989–995 (2015).
41. Raszewski, G. & Renger, S. Light harvesting in photosystem II core complexes is limited by the transfer to the trap: can the core complex turn into a photoprotective mode? *J. Am. Chem. Soc.* 130, 4431–4446 (2008).
42. Ioannidis, N. E. & Kotzabasis, K. Could structural similarity of specific domains between animal globins and plant antenna proteins provide hints important for the photoprotection mechanism? *J. Theor. Biol.* 364, 71–79 (2015).
43. van Amerongen, H. & Croce, R. Light harvesting in photosystem II. *Photosynth. Res.* 116, 251–263 (2013).
44. Berthold, D. A., Babcock, G. T. & Yocum, C. F. A highly resolved, oxygen-evolving photosystem II preparation from spinach thylakoid membranes. *FEBS Lett.* 134, 231–234 (1981).
45. Kouřil, R. *et al.* Structural characterization of a plant photosystem I and NAD(P)H dehydrogenase supercomplex. *Plant J.* 77, 568–576 (2014).
46. Sorzano, C. O. S., de la Rosa-Trevin, J. M., Tama, F. & Jonic, S. Hybrid electron microscopy normal mode analysis graphical interface and protocol. *J. Struct. Biol.* 188, 134–141 (2014).
47. Scheres, S. H. W. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* 180, 519–530 (2012).
48. Vargas, J., Álvarez-Cabrera, A.-L., Marabini, R., Carazo, J. M. & Sorzano, C. O. S. Efficient initial volume determination from electron microscopy images of single particles. *Bioinformatics* 30, 2891–2898 (2014).
49. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* 142, 334–347 (2003).
50. Rohou, A. & Grigorieff, N. CTFFIND4: fast and accurate defocus estimation from electron micrographs. *J. Struct. Biol.* 192, 216–221 (2015).
51. Li, X. *et al.* Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat. Methods* 10, 584–590 (2013).
52. Scheres, S. H. W. Beam-induced motion correction for sub-megadalton cryo-EM particles. *eLife* 3, e03665 (2014).
53. Rosenthal, P. B. & Henderson, R. Optimal determination of particle orientation, absolute hand, and contrast loss in single-particle electron cryomicroscopy. *J. Mol. Biol.* 333, 721–745 (2003).
54. Scheres, S. H. W. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nat. Methods* 9, 853–854 (2012).
55. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* 11, 63–65 (2013).
56. de la Rosa-Trevin, J. M. *et al.* Xmipp 3.0: an improved software suite for image processing in electron microscopy. *J. Struct. Biol.* 184, 321–328 (2013).
57. Pettersen, E. F. *et al.* UCSF chimera? A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612 (2004).
58. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of *coot*. *Acta Crystallogr. D* 66, 486–501 (2010).
59. Adams, P. D. *et al.* PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* 58, 1948–1954 (2002).
60. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* 68, 352–367 (2012).
61. Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N. & Sternberg, M. J. E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* 10, 845–858 (2015).
62. Touw, W. G. *et al.* A series of PDB-related databanks for everyday needs. *Nucleic Acids Res.* 43, D364–D368 (2015).
63. Chen, V. B. *et al.* Molprobity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* 66, 12–21 (2010).

## Acknowledgements

This work was supported by the FOM program ‘The thylakoid membrane—a dynamic switch (10TM02)’. S.C. is supported by the French National Research Agency Grant ANR-12-JSV8-0001-01. R.K. is supported by a Marie Curie Career Integration Grant call FP7-PEOPLE-2012-CIG and by grant LO1204 (Sustainable development of research in the Centre of the Region Haná). We acknowledge L. Franken and J. Dekker for discussions.

## Author contributions

L.S.v.B., G.T.O. and E.J.B. designed the research. S.C. isolated the supercomplex, L.S.v.B. and G.T.O. collected the data. L.S.v.B. performed the single-particle analysis. L.S.v.B., S.C., G.T.O. and A.-M.W.H.T. analysed the data. R.S.K. and R.K. analysed the negative stain supercomplex data. L.S.v.B., S.C., G.T.O., A.-M.W.H.T. and E.J.B. wrote the article.

## Additional information

Supplementary information is available for this paper.

Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints).

Correspondence and requests for materials should be addressed to E.J.B.

How to cite this article: van Bezouwen, L. S. *et al.* Subunit and chlorophyll organization of the plant photosystem II supercomplex. *Nat. Plants* 3, 17080 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Competing interests

The authors declare no competing financial interests.